Mining Periodic Behaviors for Moving Objects

Zhenhui (Jessie) Li, Bolin Ding, Jiawei Han
University of Illinois at Urbana-Champaign
Roland Kays
New York State Museum
Peter Nye
New York State Department of Environmental Conservation

SIGKDD conference
Washington DC, USA
July 28th, 2010

Work supported by NSF, ARL (NS-CTA), AFOSR (MURI), NASA, Boeing, Microsoft and IBM

Outline

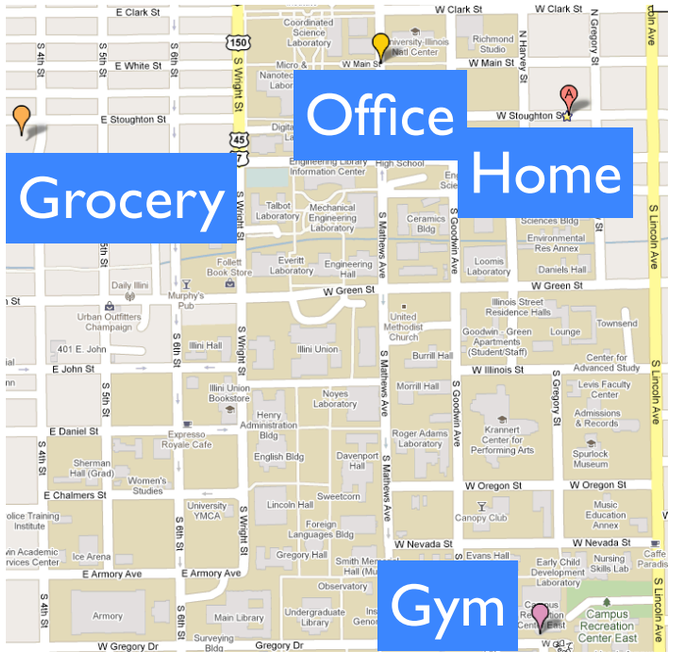
- Motivation
- Previous Method
- Periodica
- Experiment
- Summary

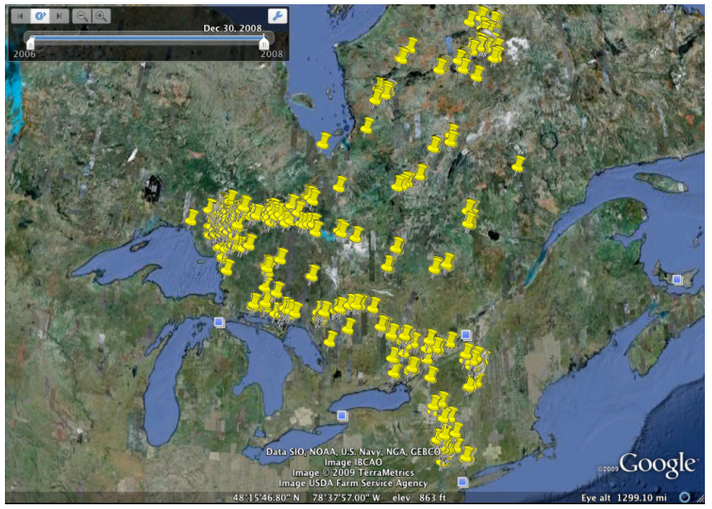
Outline

- Motivation
- Previous Method
- Periodica
- Experiment
- Summary

Periodicity is a very common phenomenon

- Moving objects usually have periodic behaviors:
  - people: go to work and go back home every weekday
  - animals: migrate yearly
- Mining periodic behaviors is useful to:
  - summarize over long historical movement
  - predict future movement
  - detect abnormal event





my periodic behavior:
10:00am office
1:00pm home
2:00pm office
6:00pm home
7:30pm office
11:00pm home

gym, tues. & thurs.
grocery, weekend

bald eagle:
yearly migration

The problem of mining periodic behavior

- Given movement sequence with constant time gap (i.e., one point every hour) of one moving object, $LOC = loc_1 loc_2 \cdots loc_n$
- Goal: (1) detect periods and (2) summarize corresponding periodic behaviors

	8:00	9:00	10:00	$\cdots$	17:00	18:00	19:00
dorm	0.9	0.2	0.1	$\cdots$	0.2	0.7	0.8
office	0.05	0.7	0.95	$\cdots$	0.75	0.2	0.1
unknown	0.05	0.1	0.05	$\cdots$	0.05	0.1	0.1

a possible periodic behavior for daily period

Mining multiple interleaving periodic behaviors is challenging

Raw data of David's movement

...
2009–02–05 07:01 (601, 254)
2009–02–05 09:14 (811, 60)
2009–02–05 10:58 (810, 55)
2009–02–05 14:29 (820, 100)
...
2009–06–12 09:56 (110, 98)
2009–06–12 11:20 (101, 65)
2009–06–12 20:08 (20, 97)
2009–06–12 22:19 (15, 100)
...

Periodic behaviors

- Periodic Behavior #1
  (Period: day; Time span: Sept. – May)
  9:00–18:00 in the office
  20:00–8:00 in the dorm
- Periodic Behavior #2
  (Period: day; Time span: June – Aug.)
  8:00–18:00 in the company
  20:00–7:30 in the apartment
- Periodic Behavior #3
  (Period: week; Time span: Sept. – May)
  13:00–15:00 Mon. and Wed. in the classroom
  14:00–16:00 Tues. and Thurs. in the gym

different times

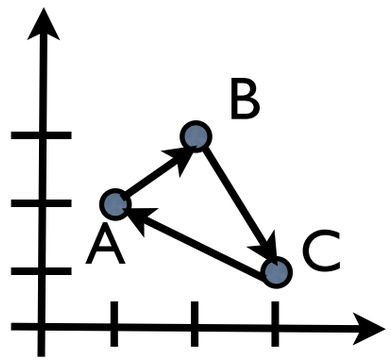
multiple periods

different locations

Outline

- Motivation
- Previous Method
- Periodica
- Experiment
- Summary

Directly applying FFT on the movement sequence



movement sequence:
(1,2), (2,3), (3,1), (1,2), (2,3), (3,1) ...

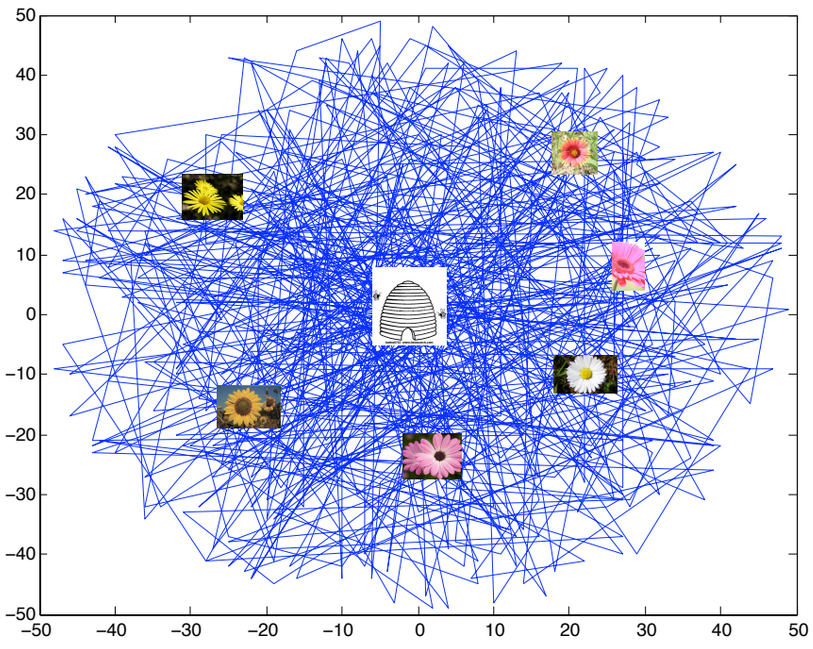
- Transform points into complex plane

(x,y) ⟶ x+yi
1+2i 2+3i 3+1i 1+2i 2+3i 3+1i ...

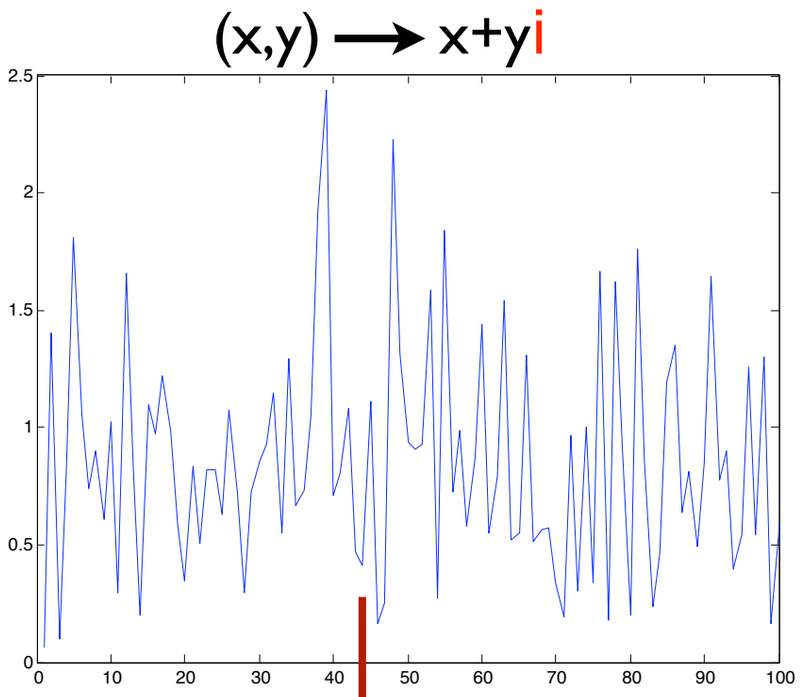
(x,y) ⟶ y+xi
2+1i 3+2i 1+3i 2+1i 3+2i 1+3i ...

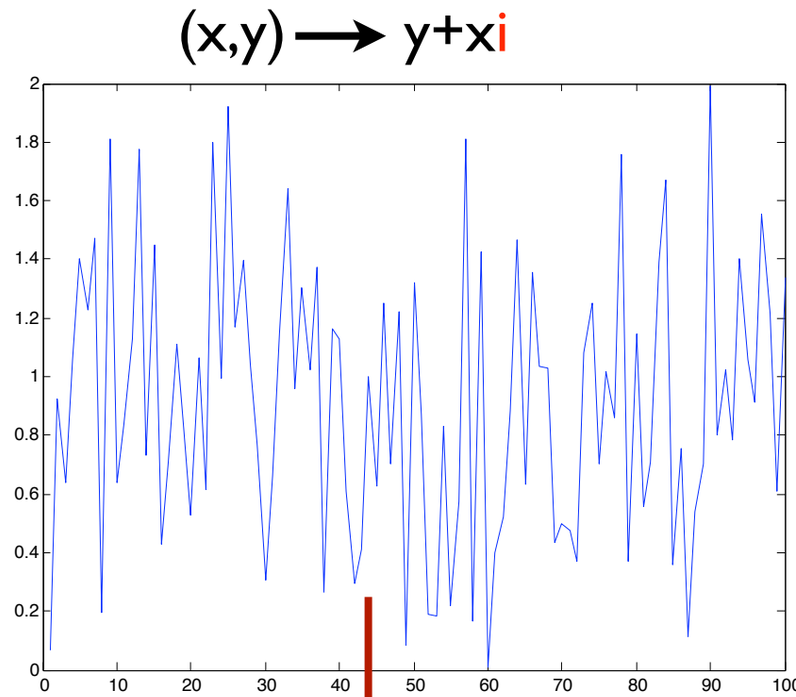
- Apply Discrete Fourier Transform on complex plane

Previous method is sensitive to trajectory noise



Bee example:
8 hours in hive
16 hours fly nearby

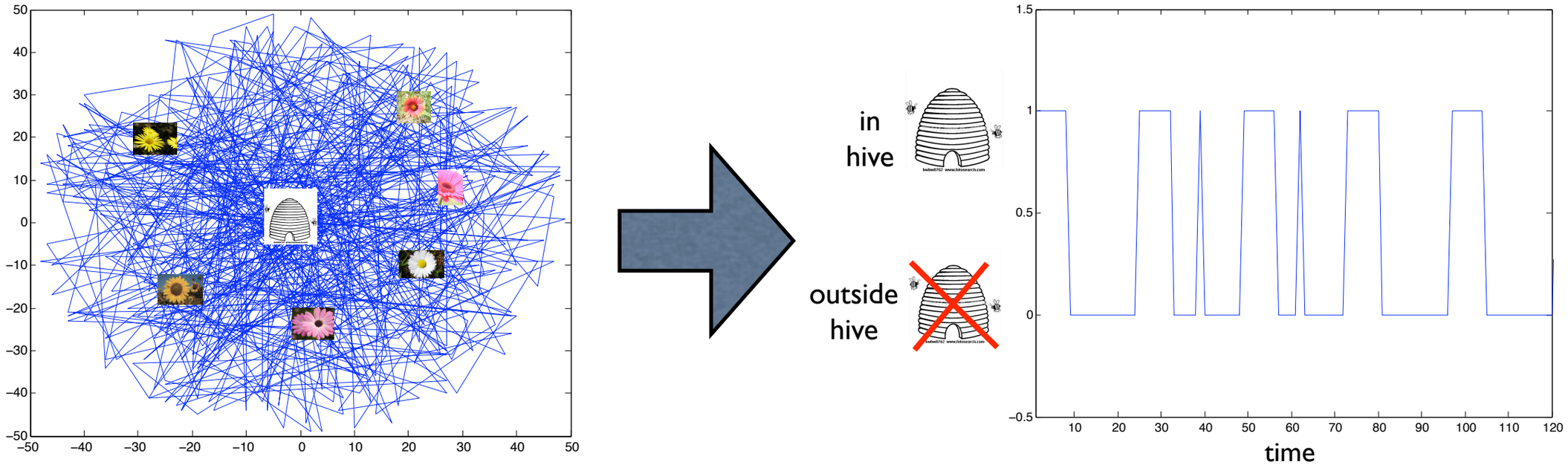




Method fails to detect periods.
It should have strongest power at **42.7**
(T = 24, NFFT/T = 1024/24 = 42.7).

Find the right spot to observer the movement

- The concrete trajectory is not important.
- We can observe its movement from the hive (in or out).

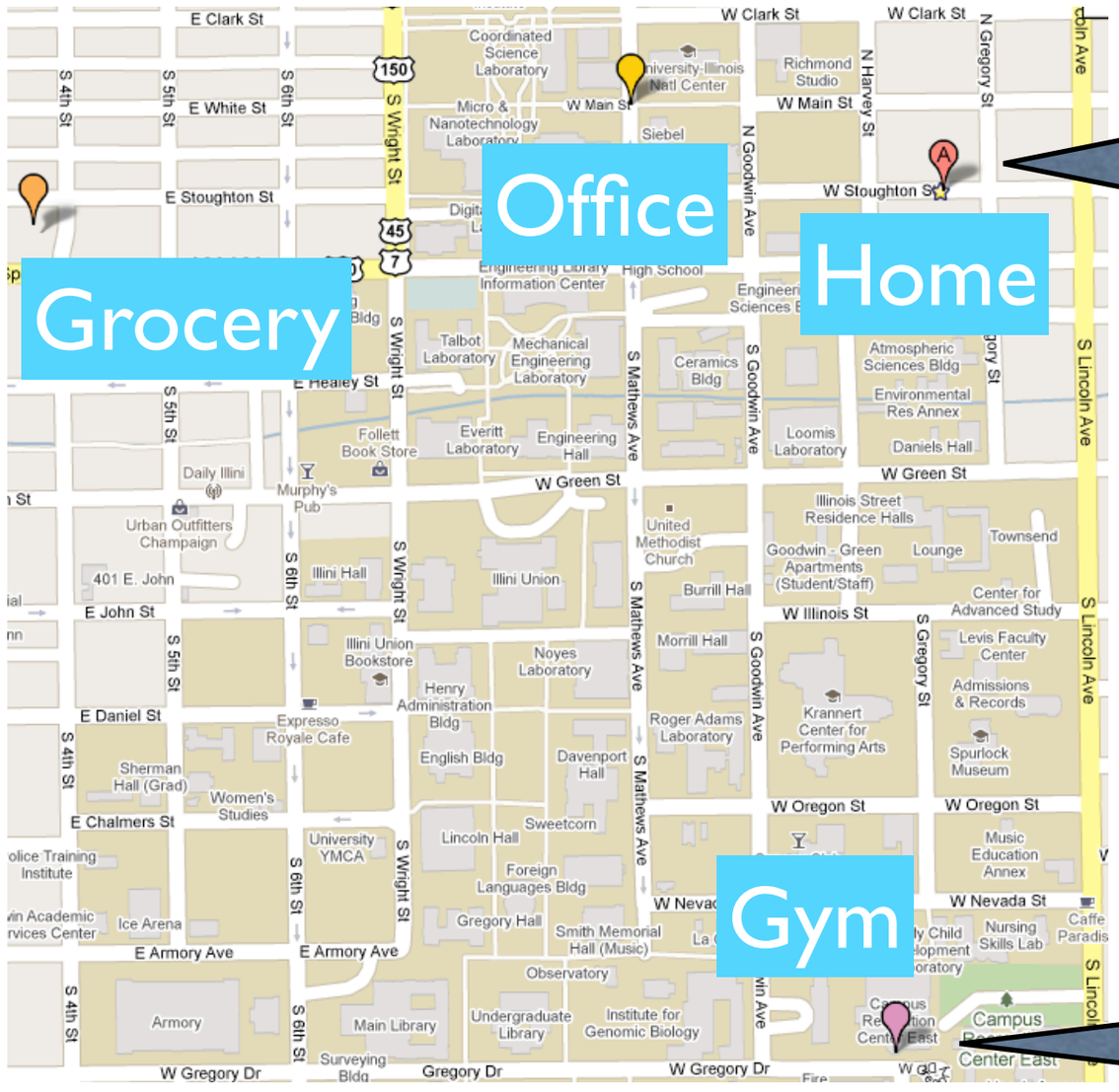


The movement is transformed into a binary sequence (in hive or outside hive).

The period in the binary sequence is easy to be detected.

reference spots = frequently visited locations

By observing movement from reference spots, the periods are easier to be detected.



reference spots = frequently visited locations

By observing movement from reference spots, the periods are easier to be detected.

Key point: observe the movement from some reference spots

observe my movement from my home (at home or not at home) -- daily period can be detected

observe my movement from the gym (at gym or not at gym) -- weekly period can be detected

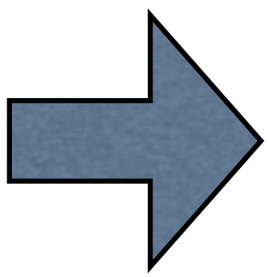
Outline

- Motivation
- Previous Method
- Periodica
- Experiment
- Summary

Periodica outline

- Step 1: Detect periods
  - find reference spots
  - for each reference spot:
    - movement is transformed into a binary sequence
    - detect periods in the binary sequence
- Step 2: Summarize periodic behaviors
  - for each period, segment the movement by period
  - hierarchically cluster segments
  - a behavior is summarized over the segments in a cluster

Periodica outline



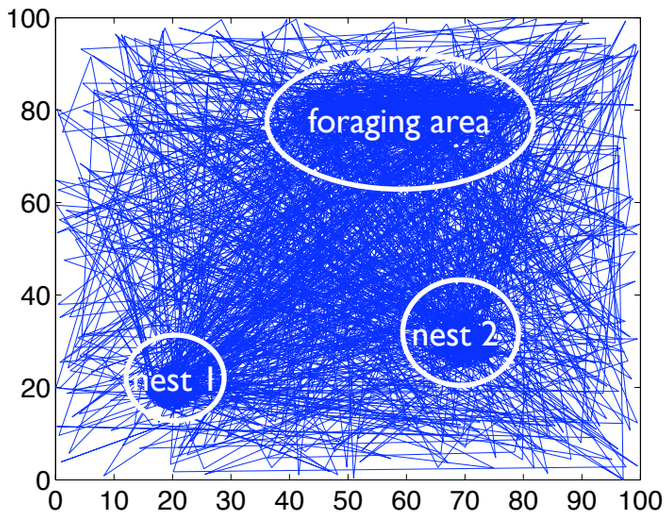
Step 1: Detect periods

- find reference spots
- for each reference spot:
  - movement is transformed into a binary sequence
  - detect periods in the binary sequence

- Step 2: Summarize periodic behaviors
  - for each period, segment the movement by period
  - hierarchically cluster segments
  - a behavior is summarized over the segments in a cluster

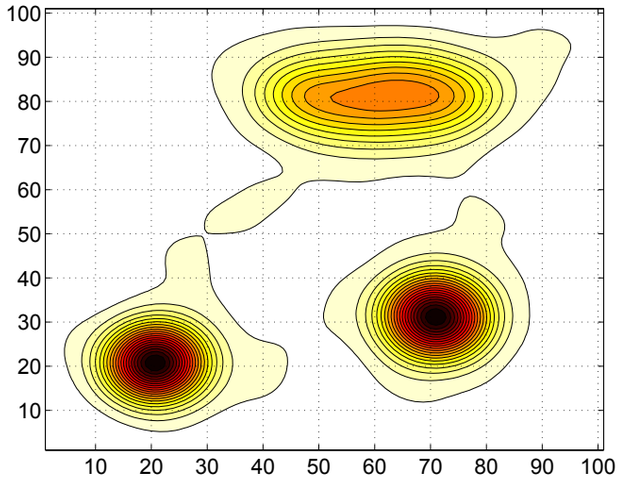
Periodica: Detect periods: find reference spots

Running Example

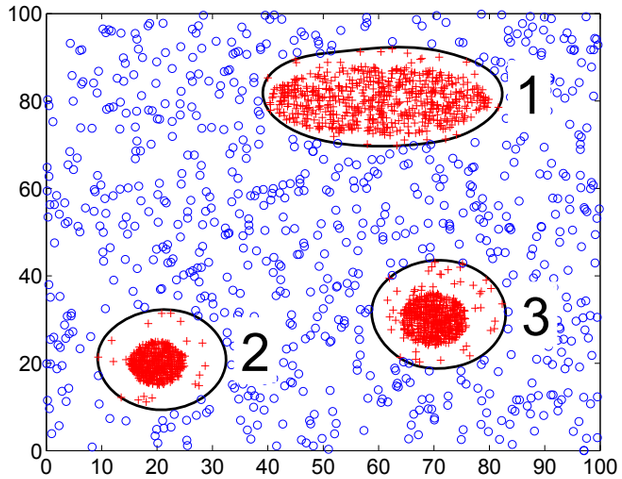


- first 50 days: daily periodic behavior between nest and foraging area
- second 50 days: daily periodic behavior between **another** nest and **the same** foraging area

Reference spot:
(1) frequently visited regions/locations;
(2) higher density than a random location



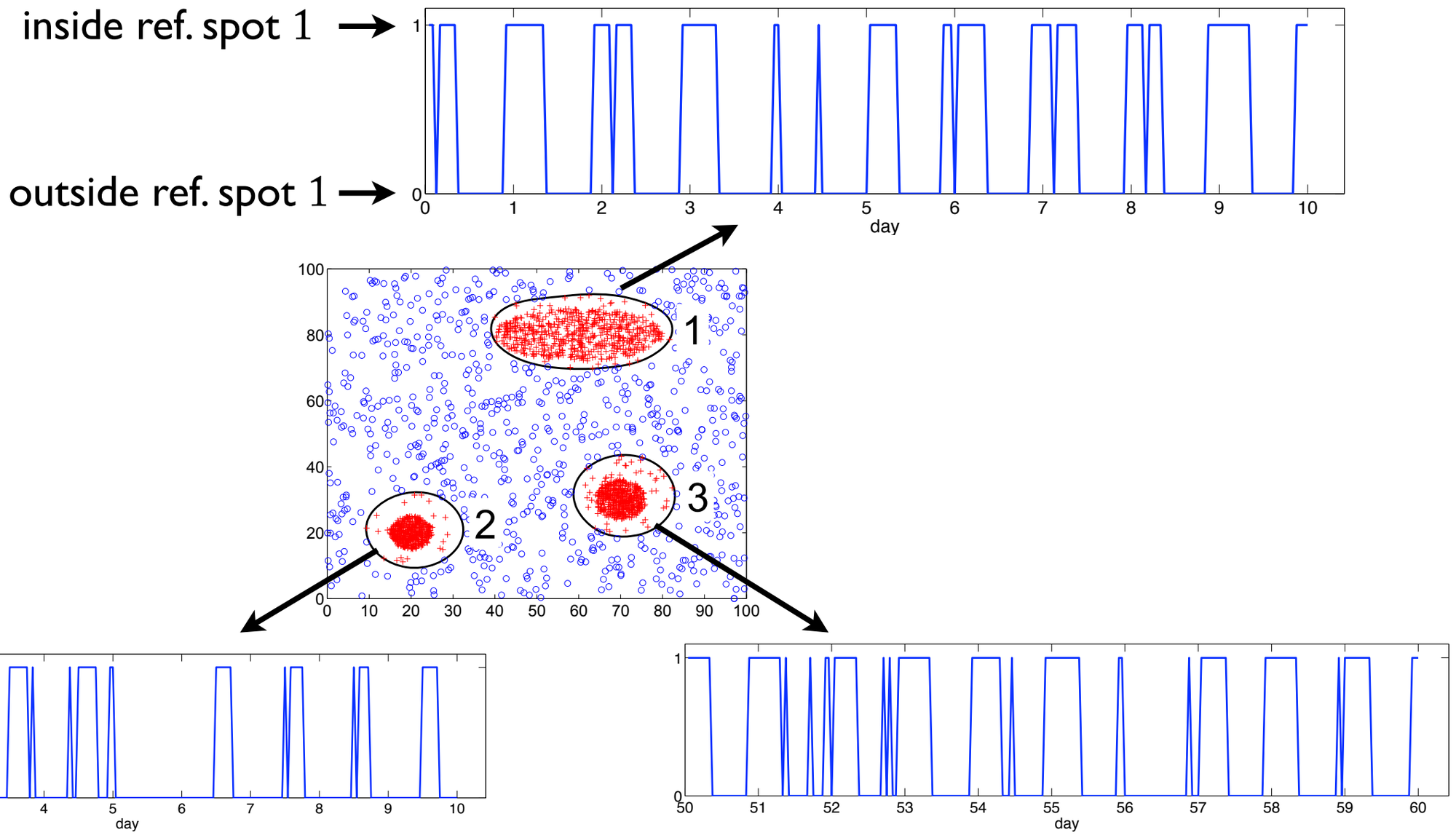
Use kernel-based method to calculate the densities



Reference spots: contours of high density places

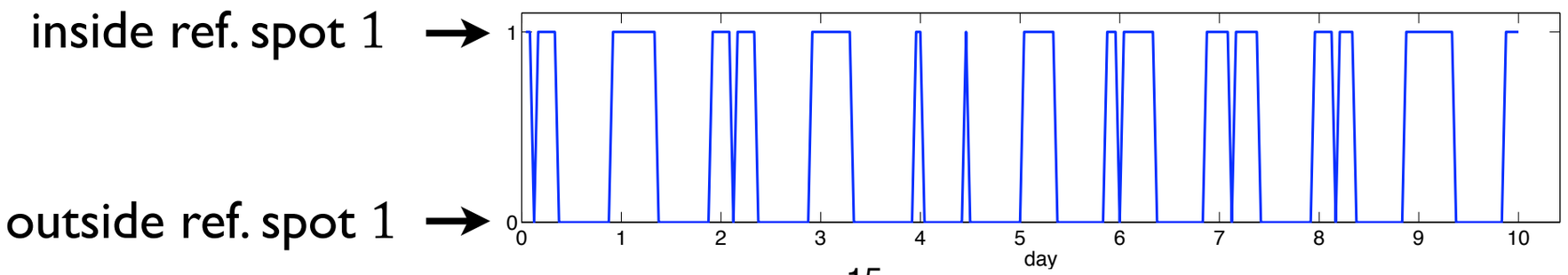
Periodica: Detect periods:

transform into in-and-out binary sequence

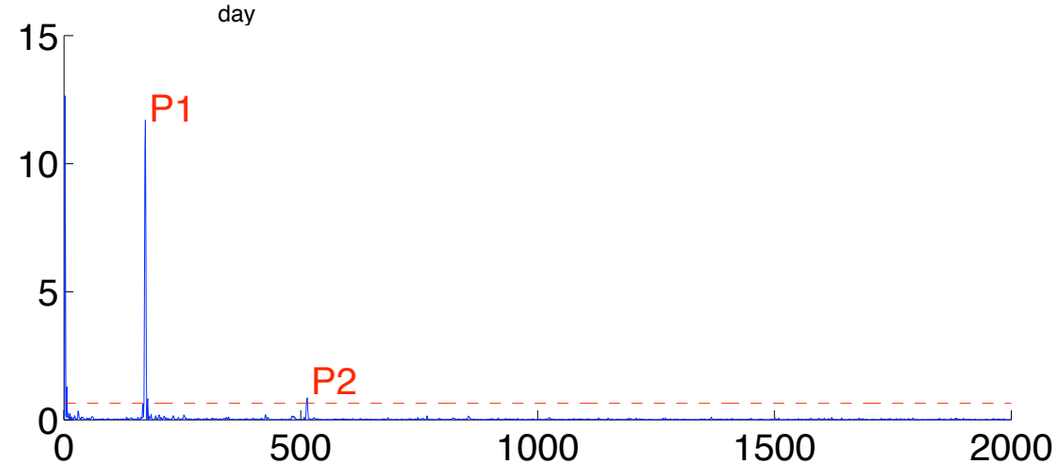


Periodica: Detect periods:

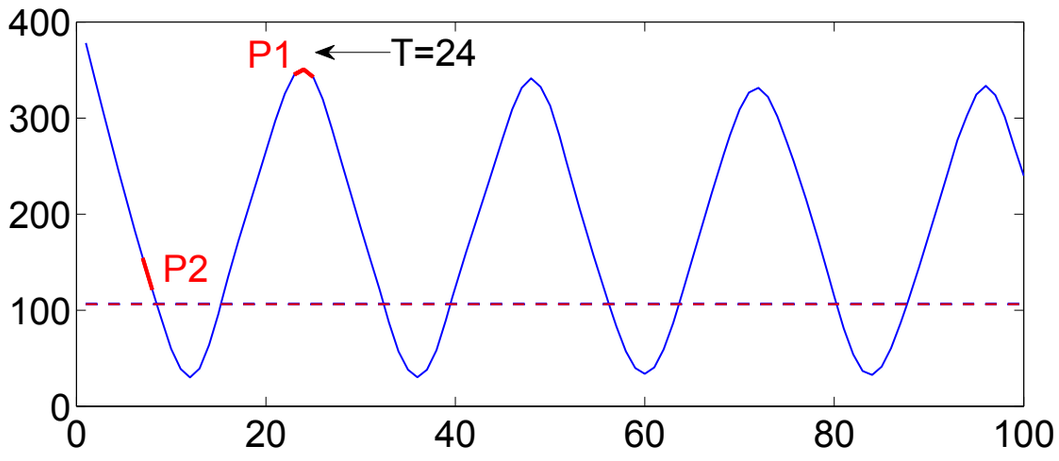
detect periods in binary sequence



- Fourier transform (periodogram) will give a range of periods.
- Autocorrelation further confirms the exact periods.

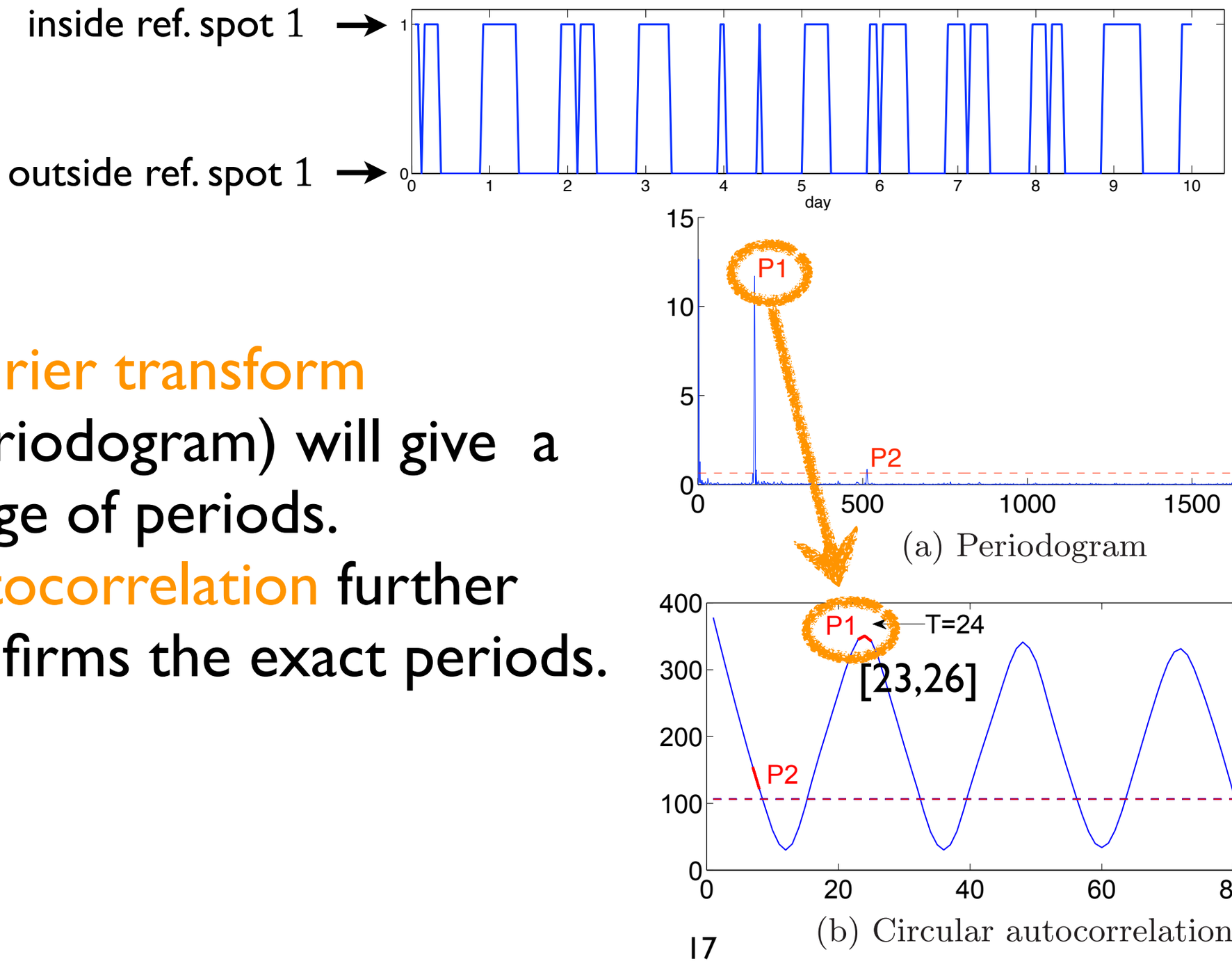


(a) Periodogram



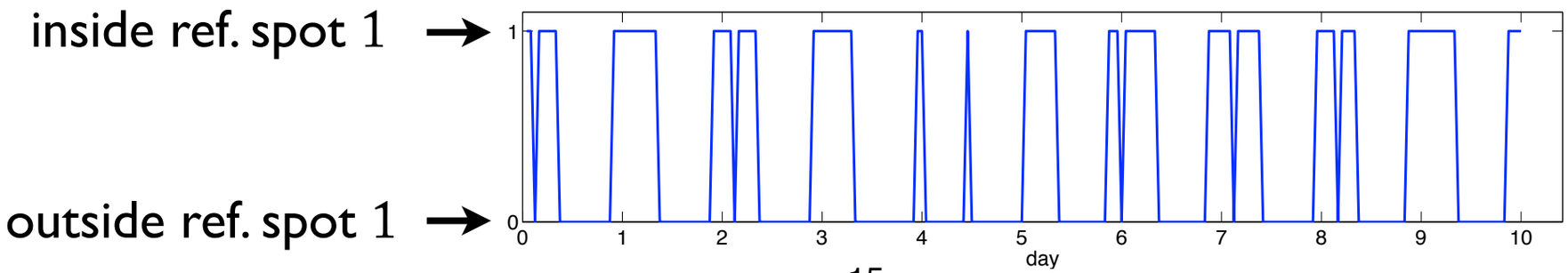
(b) Circular autocorrelation

Periodica: Detect periods:
detect periods in binary sequence

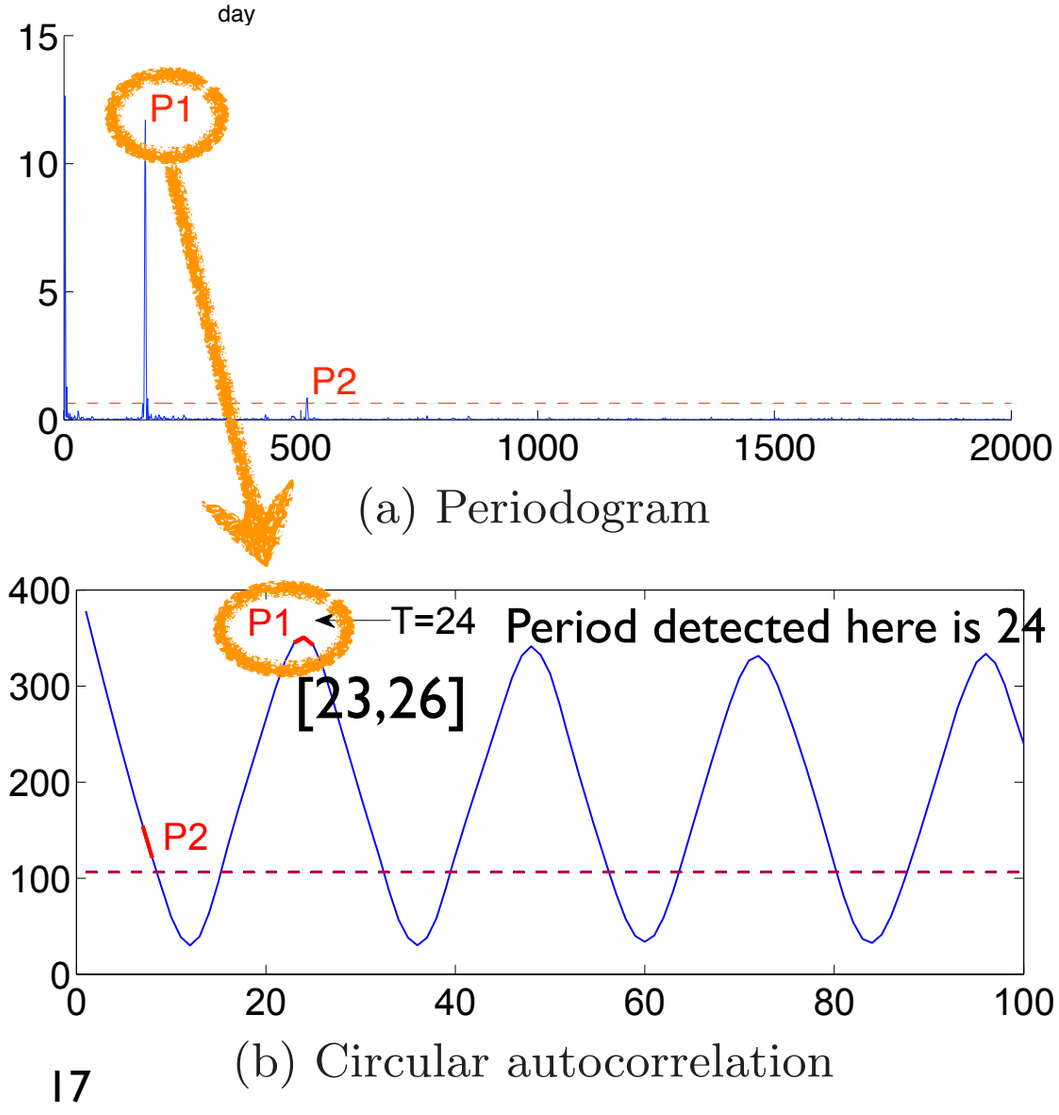


- Fourier transform (periodogram) will give a range of periods.
- Autocorrelation further confirms the exact periods.

Periodica: Detect periods:
detect periods in binary sequence

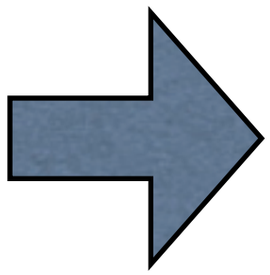


- Fourier transform (periodogram) will give a range of periods.
- Autocorrelation further confirms the exact periods.



Periodica outline

- Step 1: Detect periods
  - find reference spots
  - for each reference spot:
    - movement is transformed into a binary sequence
    - detect periods in the binary sequence



Step 2: Summarize periodic behaviors

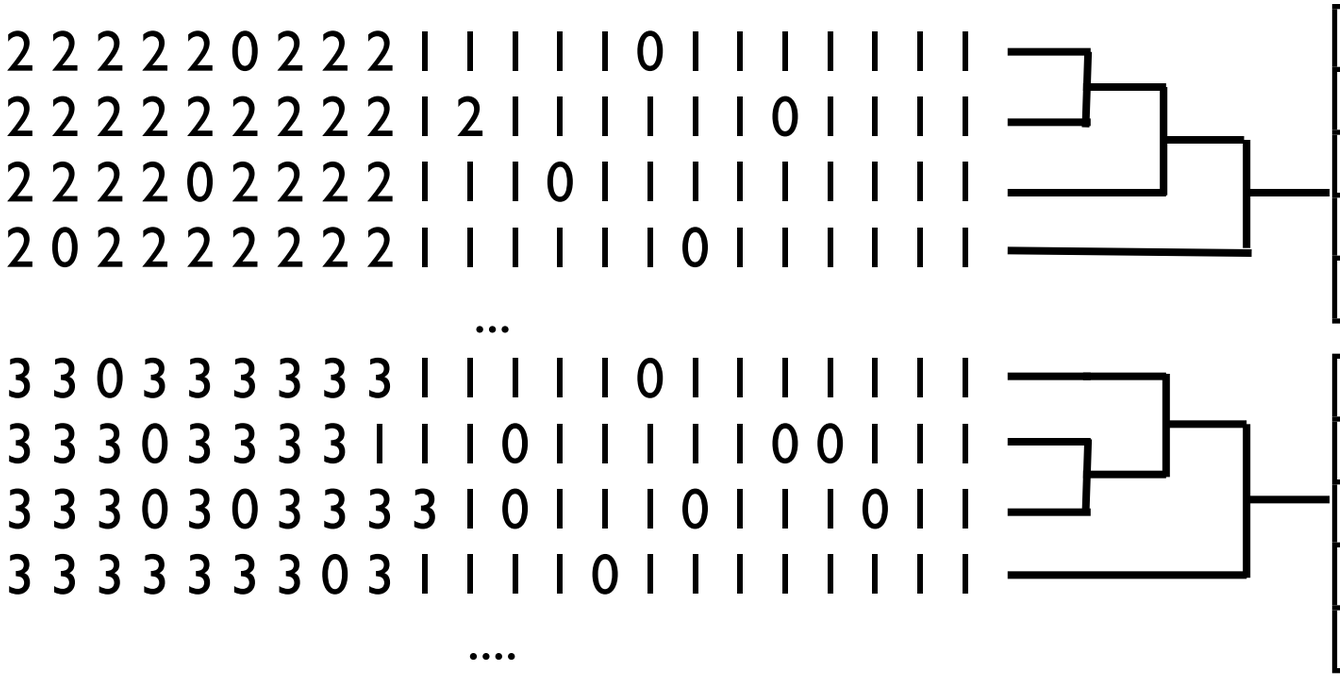
- for each period, segment the movement by period
- hierarchically cluster segments
- a behavior is summarized over the segments in a cluster

Periodica: Summarize behaviors: segment movements using the period

- First, the movement is symbolized using ref. spots. (0 means it is outside any ref. spot.)
- Given the period T=24(hours), the movement is segmented into "day"s.

day 1	2 2 2 2 2 0 2 2 2 1 1 1 1 1 0 1 1 1 1 1 1 1
day 2	2 2 2 2 2 2 2 2 2 1 2 1 1 1 1 1 1 0 1 1 1 1
	…
day n	3 3 0 3 3 3 3 3 3 1 1 1 1 1 0 1 1 1 1 1 1 1
day n+1	3 3 0 3 3 3 3 3 1 1 1 0 1 1 1 1 1 0 0 1 1 1
	….

Periodica: Summarize behaviors: hierarchically cluster segments



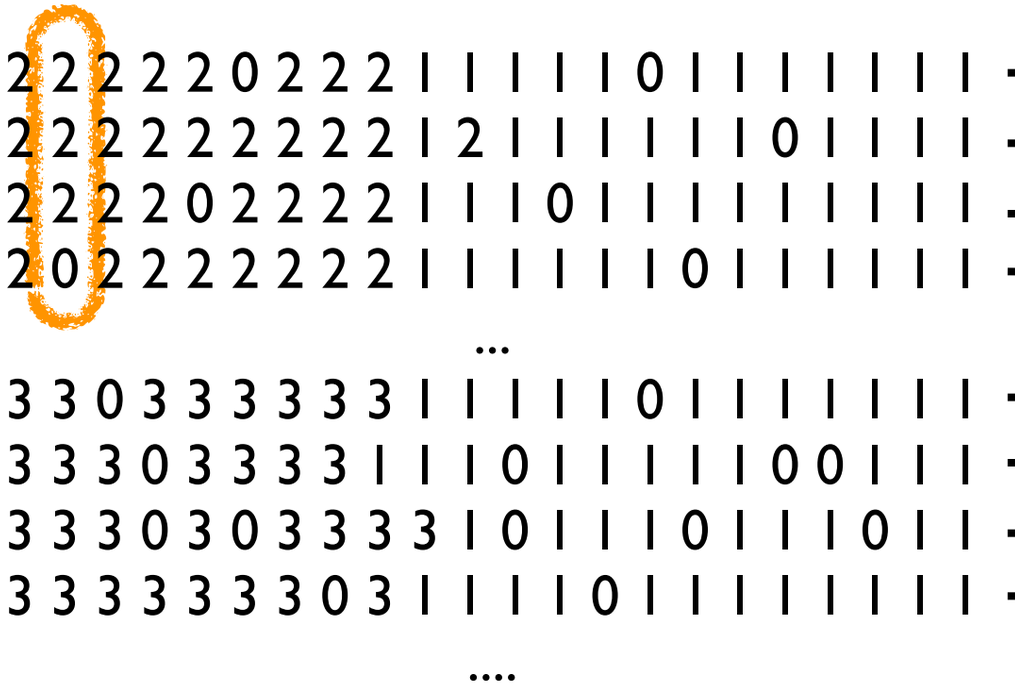
spot	1	2	3	4	...	21	22	23	24
1	0	0	0	0	...	0.75	1.0	1.0	1.0
2	1.0	0.75	1.0	1.0	...	0	0	0	0
3	0	0	0	0	...	0	0	0	0
0	0	0.25	0	0	...	0.25	0	0	0

spot	1	2	3	4	...	21	22	23	24
1	0	0	0	0	...	0.75	0.75	1.0	1.0
2	0	0	0	0	...	0	0	0	0
3	1.0	1.0	0.75	0.5	...	0	0	0	0
0	0	0	0.25	0	...	0.25	0.25	0	0

cluster (a set of segments) = behavior = probability matrix

- Bottom-up hierarchical clustering.
- Initially, each segment is a behavior.
- The distance between behaviors are calculated using KL-divergence.

Periodica: Summarize behaviors: hierarchically cluster segments



spot	1	2	3	4	...	21	22	23	24
1	0	0	0	0	...	0.75	1.0	1.0	1.0
2	1.0	0.75	1.0	1.0	...	0	0	0	0
3	0	0	0	0	...	0	0	0	0
0	0	0.25	0	0	...	0.25	0	0	0

spot	1	2	3	4	...	21	22	23	24
1	0	0	0	0	...	0.75	0.75	1.0	1.0
2	0	0	0	0	...	0	0	0	0
3	1.0	1.0	0.75	0.5	...	0	0	0	0
0	0	0	0.25	0	...	0.25	0.25	0	0

cluster (a set of segments) = behavior = probability matrix

- Bottom-up hierarchical clustering.
- Initially, each segment is a behavior.
- The distance between behaviors are calculated using KL-divergence.

Periodica: Summarize behaviors: hierarchically cluster segments

spot	1	2	3	4	...	21	22	23	24
1	0	0	0	0	...	0.75	1.0	1.0	1.0
2	1.0	0.75	1.0	1.0	...	0	0	0	0
3	0	0	0	0	...	0	0	0	0
0	0	0.25	0	0	...	0.25	0	0	0

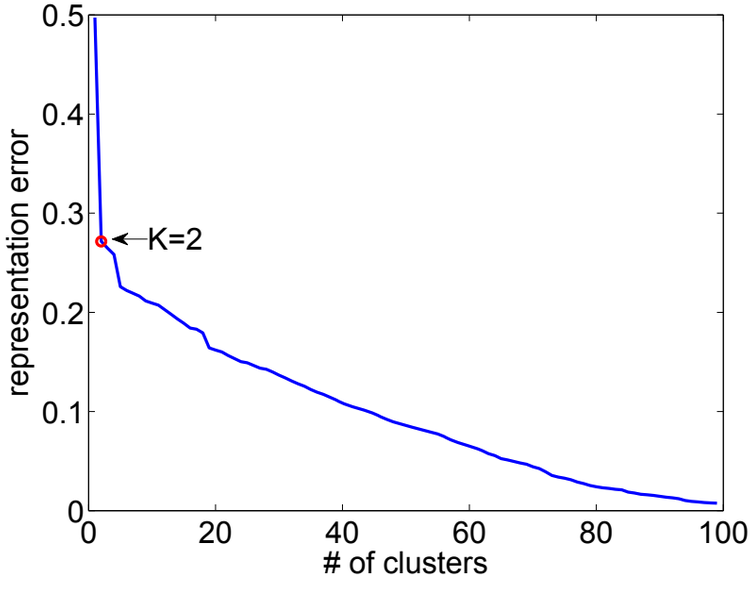
spot	1	2	3	4	...	21	22	23	24
1	0	0	0	0	...	0.75	0.75	1.0	1.0
2	0	0	0	0	...	0	0	0	0
3	1.0	1.0	0.75	0.5	...	0	0	0	0
0	0	0	0.25	0	...	0.25	0.25	0	0

cluster (a set of segments) = behavior = probability matrix

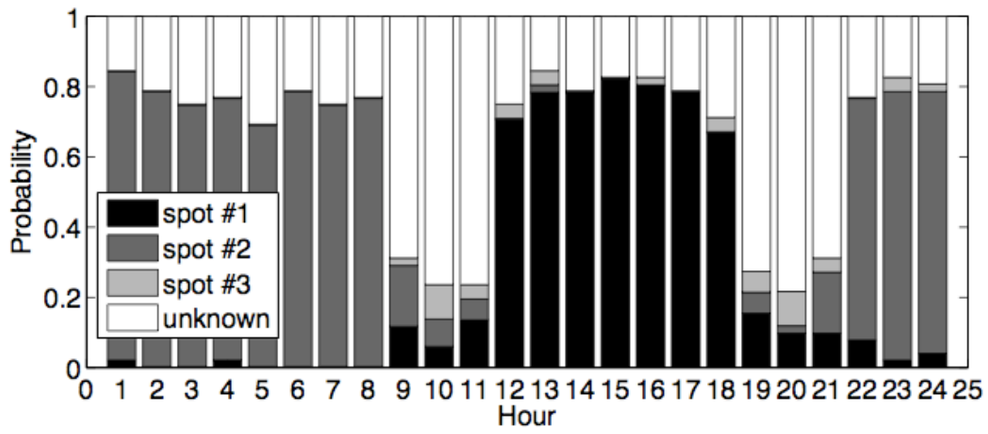
- Bottom-up hierarchical clustering.
- Initially, each segment is a behavior.
- The distance between behaviors are calculated using KL-divergence.

Periodica: Summarize behaviors: the number of periodic behaviors

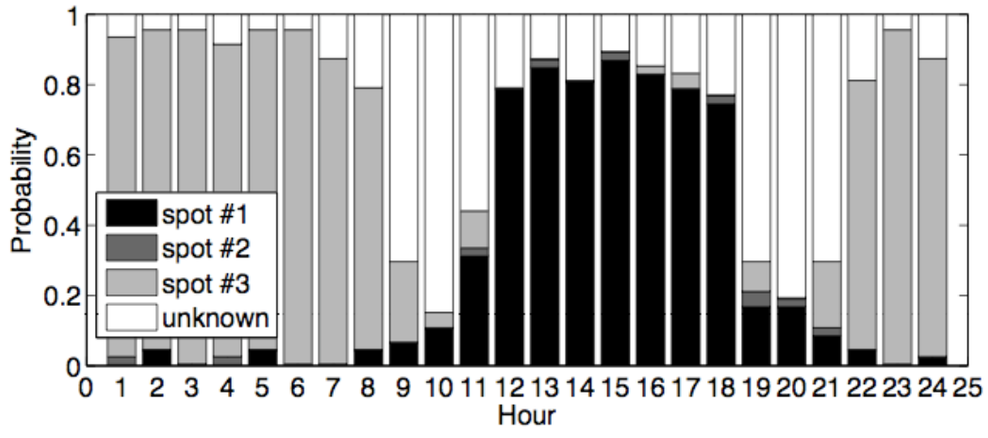
- Use representation error to monitor the "purity" of segments in a cluster.
- Rep. error increases as clusters merge.
- The sharp increase indicates the right number of periodic behaviors.



Finally, two periodic behaviors are detected.



(a) $\mathbf{P}$ of periodic behavior #1



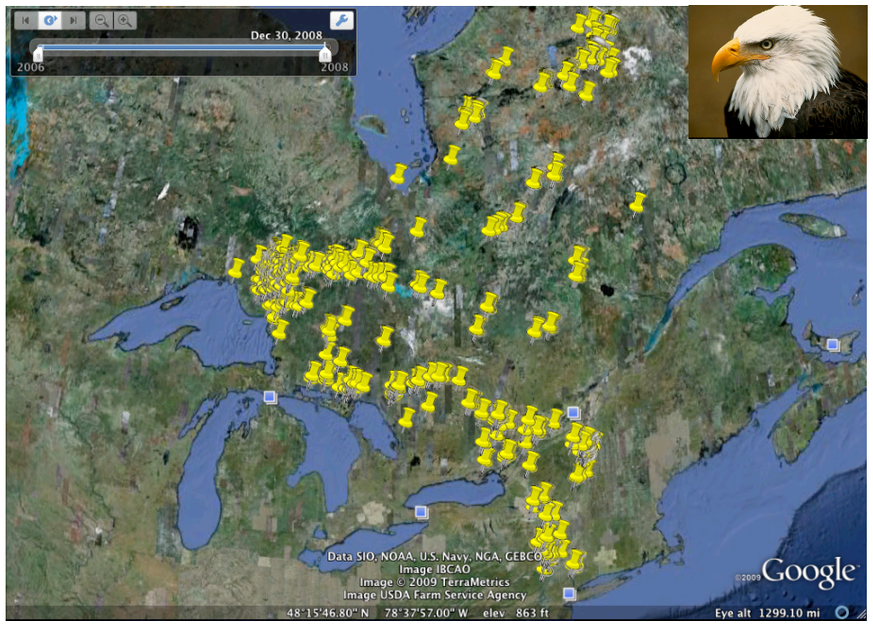
(b) $\mathbf{P}$ of periodic behavior #2

Outline

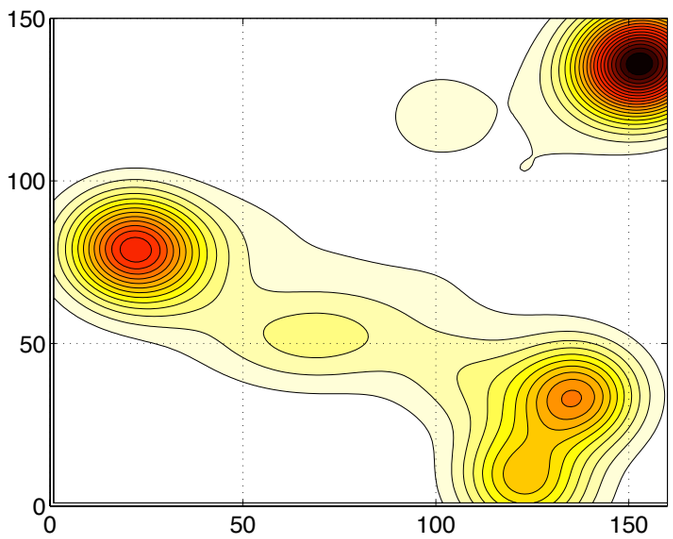
- Motivation
- Previous Method
- Periodica
- Experiment
- Summary

A real bald eagle:

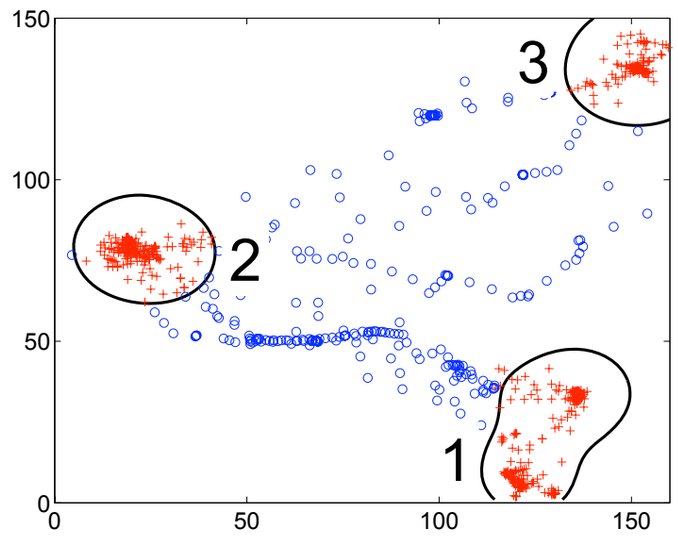
Step 1 Detect periods



Original movement data:
Time span: Jan, 2006 to Dec., 2008
Number of points recorded: 2204
Movement is linearly interpolated using time gap as 24 hours



density map



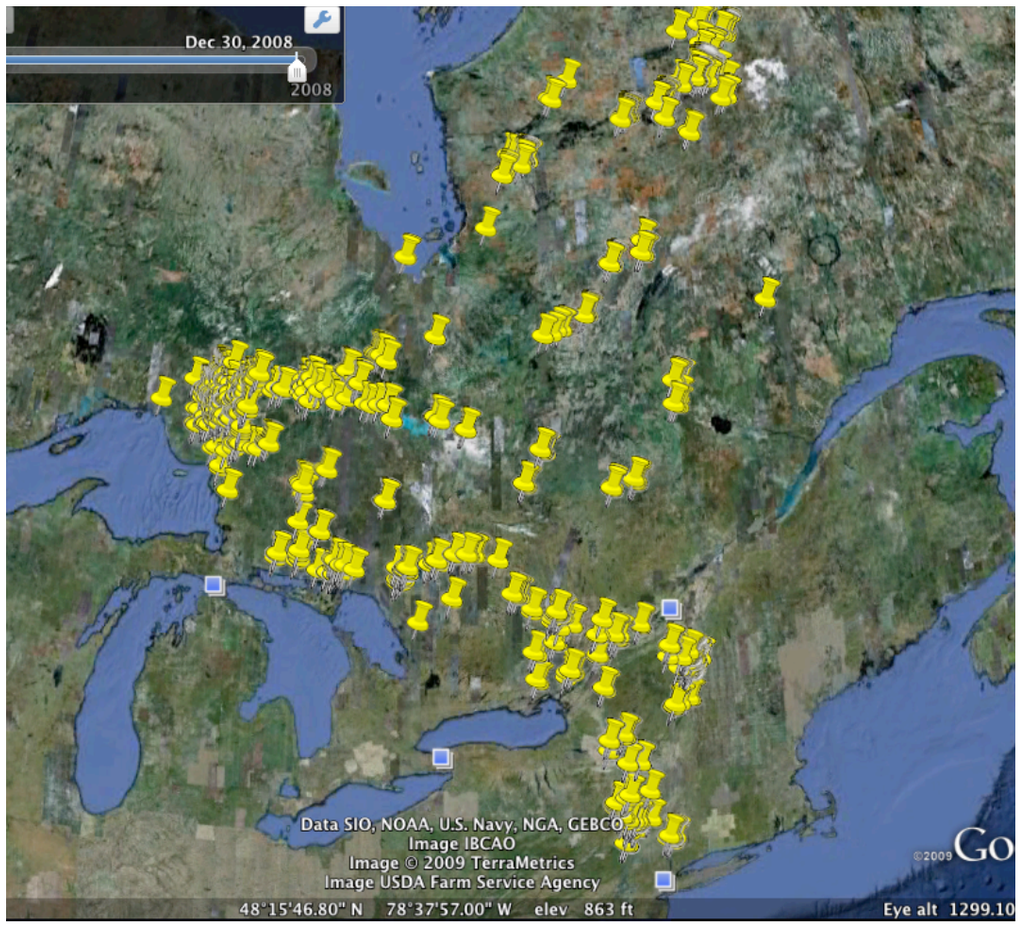
reference spots

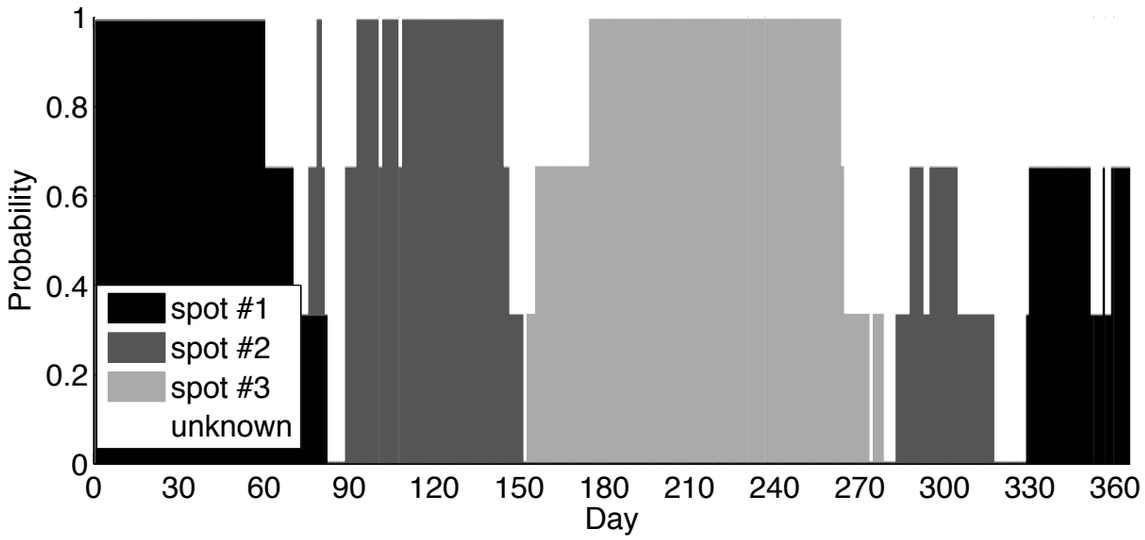
Period detected using Fourier transform and autocorrelation:

ref. spot	period
1	363
2	363
3	364

A real bald eagle:
Step 2 Summarize behaviors

- Segment movement by 363 days
- "year"s are clustered into one cluster

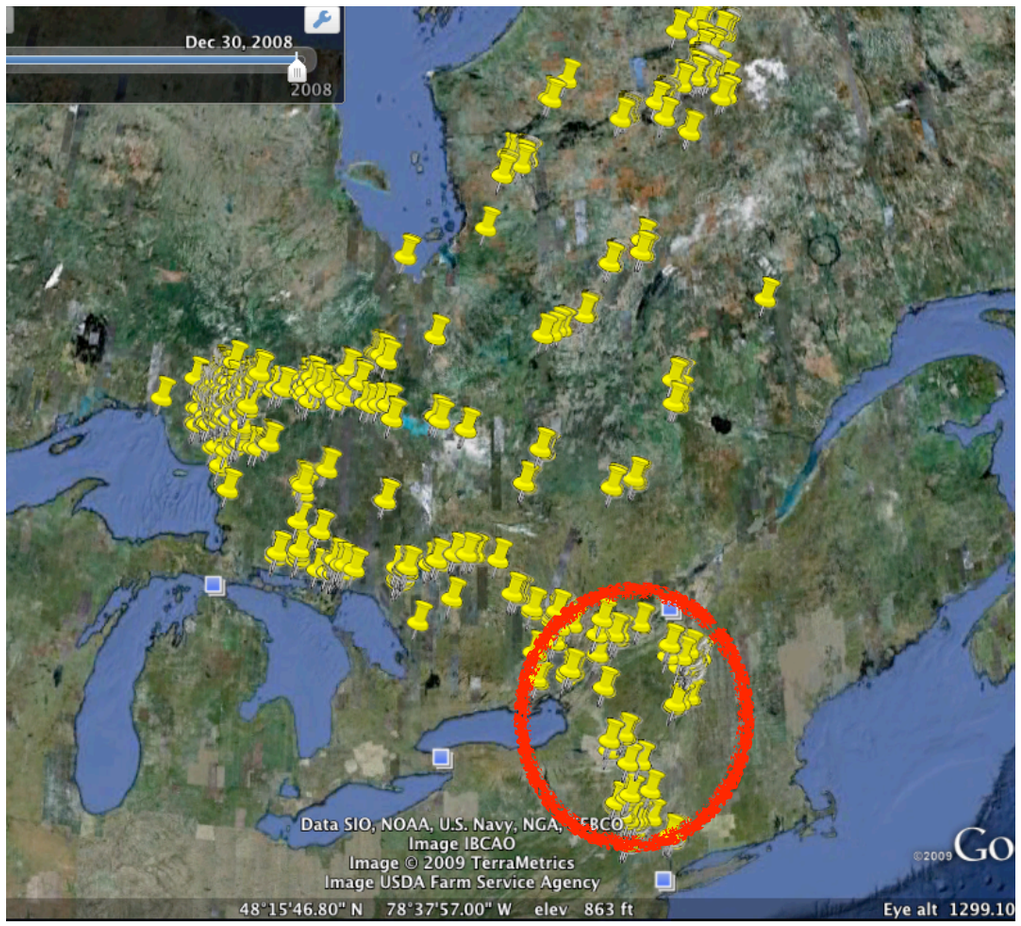


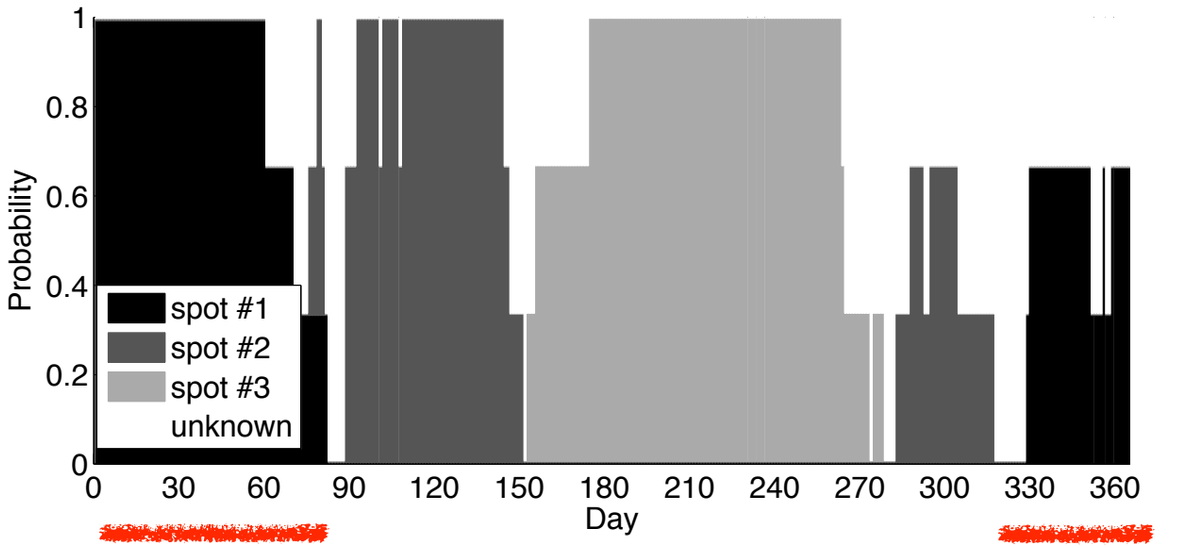


"This bald eagle stays in New York area (i.e., reference spot # 1) from December to March. In March, it flies to Great Lakes area (i.e., reference spot #2) and stays there until the end of May. It flies to Quebec area (i.e., reference spot #3) in the summer and stays there until late September. Then it flies back to Great Lake again staying there from mid October to mid November and goes back to New York in December."

A real bald eagle:
Step 2 Summarize behaviors

- Segment movement by 363 days
- "year"s are clustered into one cluster



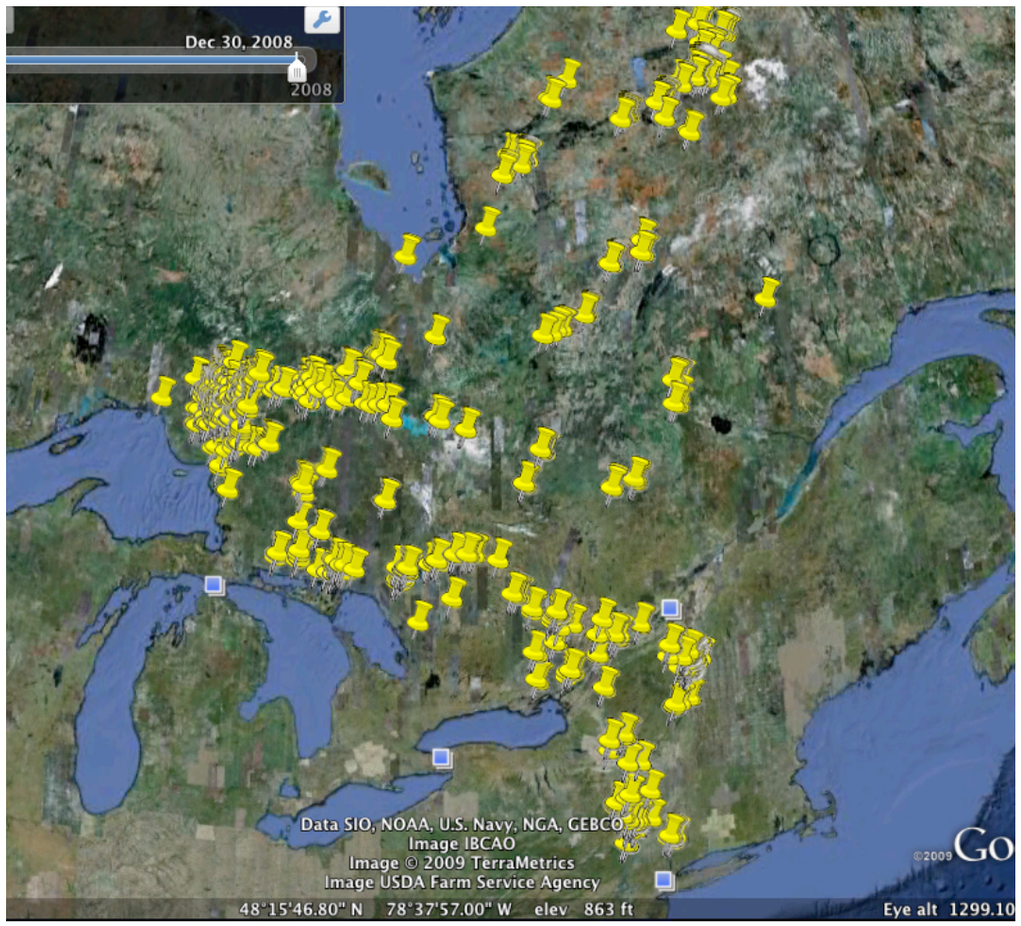


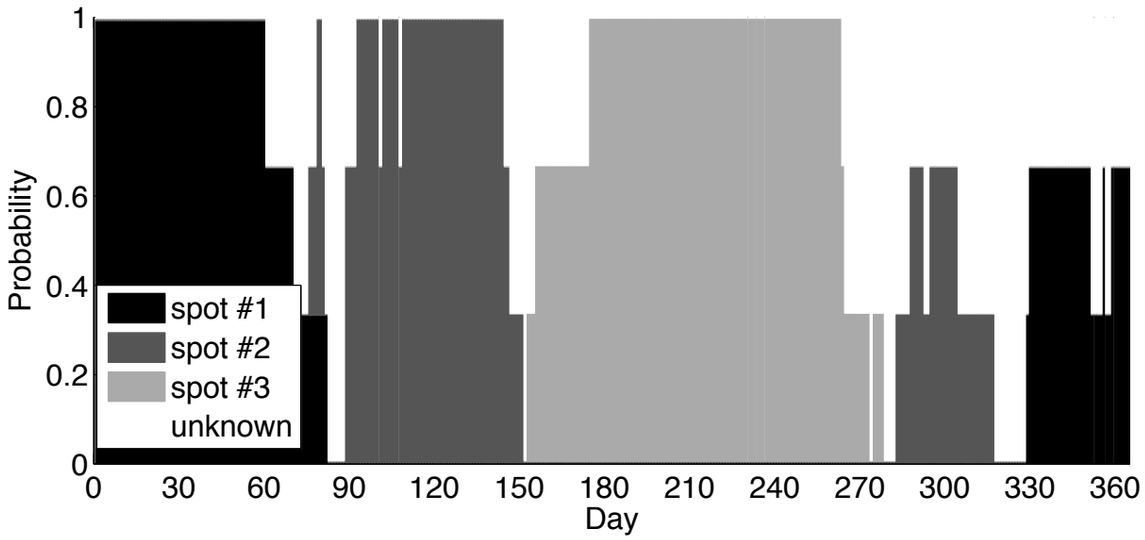
March December

"This bald eagle stays in New York area (i.e., reference spot # 1) from December to March. In March, it flies to Great Lakes area (i.e., reference spot #2) and stays there until the end of May. It flies to Quebec area (i.e., reference spot #3) in the summer and stays there until late September. Then it flies back to Great Lake again staying there from mid October to mid November and goes back to New York in December."

A real bald eagle:
Step 2 Summarize behaviors

- Segment movement by 363 days
- "year"s are clustered into one cluster



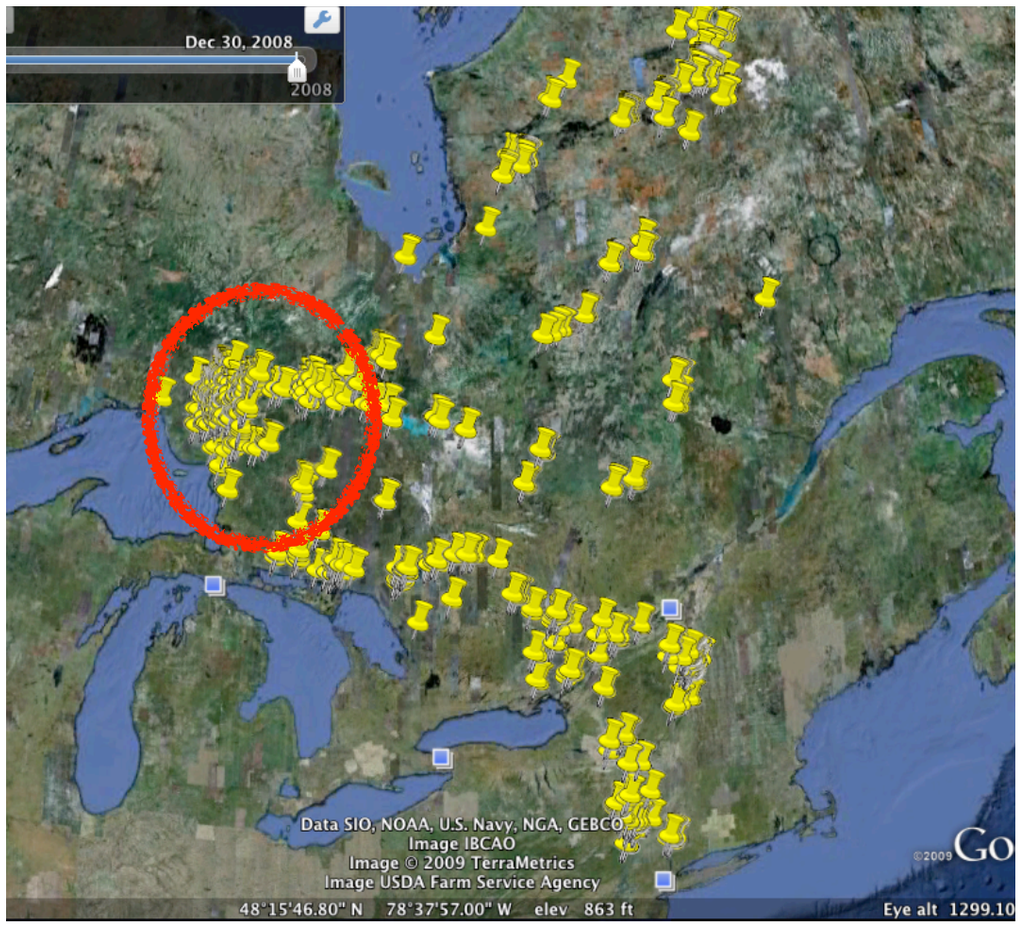


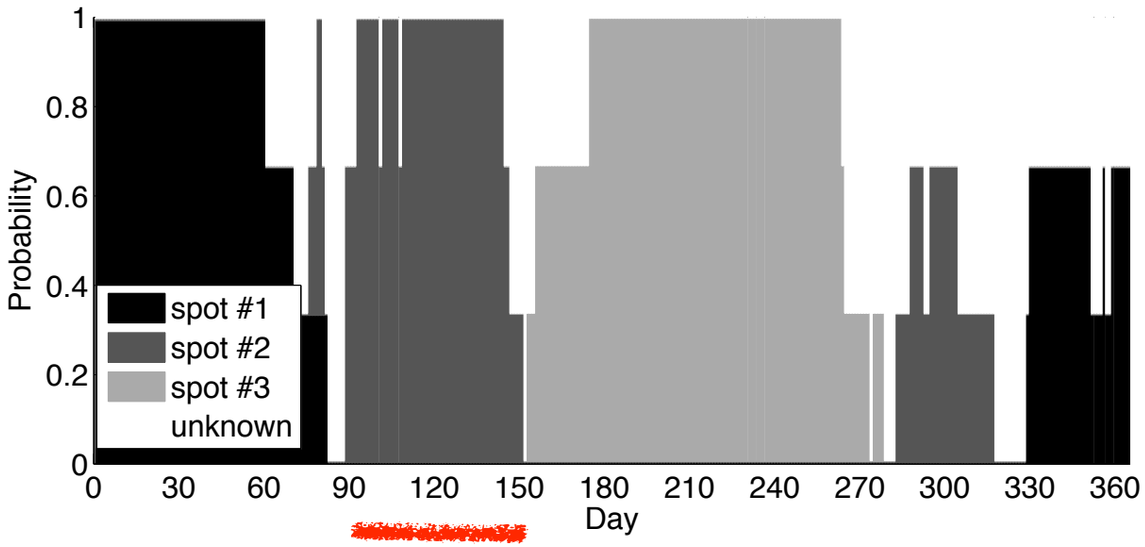
March December

"This bald eagle stays in New York area (i.e., reference spot # 1) from December to March. In March, it flies to Great Lakes area (i.e., reference spot #2) and stays there until the end of May. It flies to Quebec area (i.e., reference spot #3) in the summer and stays there until late September. Then it flies back to Great Lake again staying there from mid October to mid November and goes back to New York in December."

A real bald eagle:
Step 2 Summarize behaviors

- Segment movement by 363 days
- "year"s are clustered into one cluster



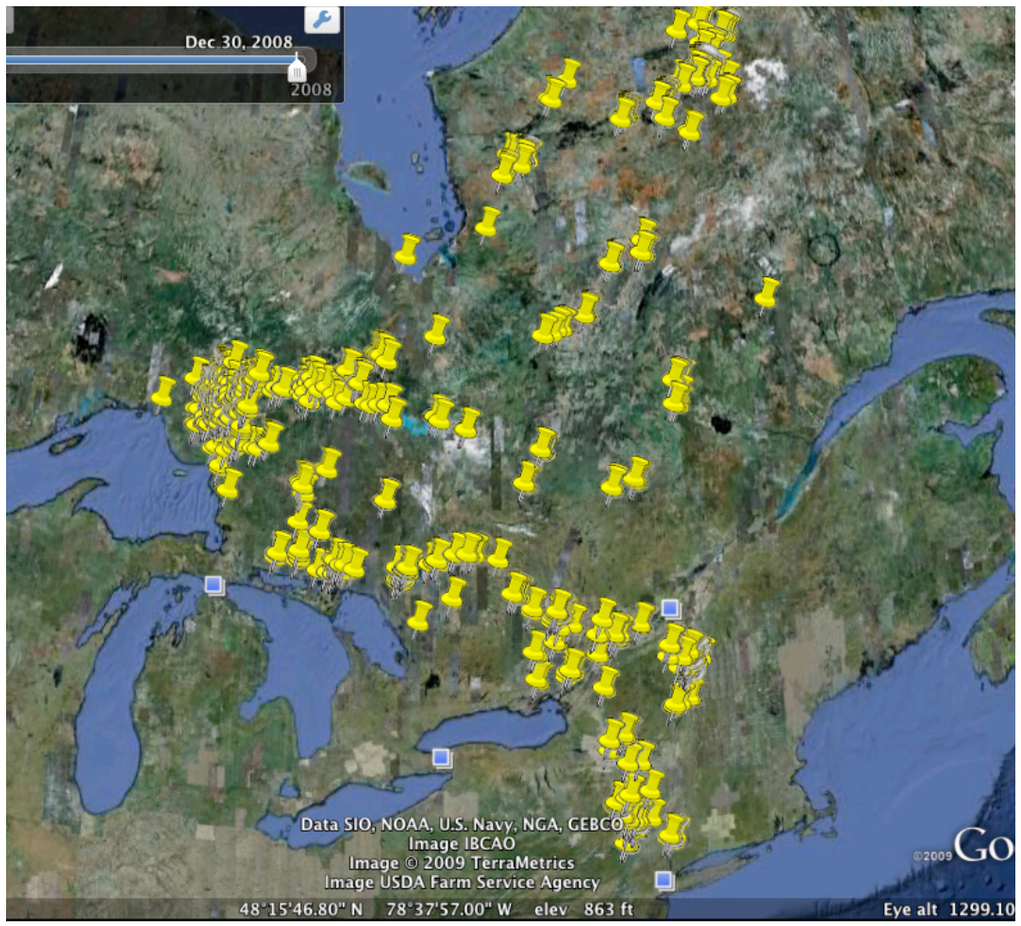


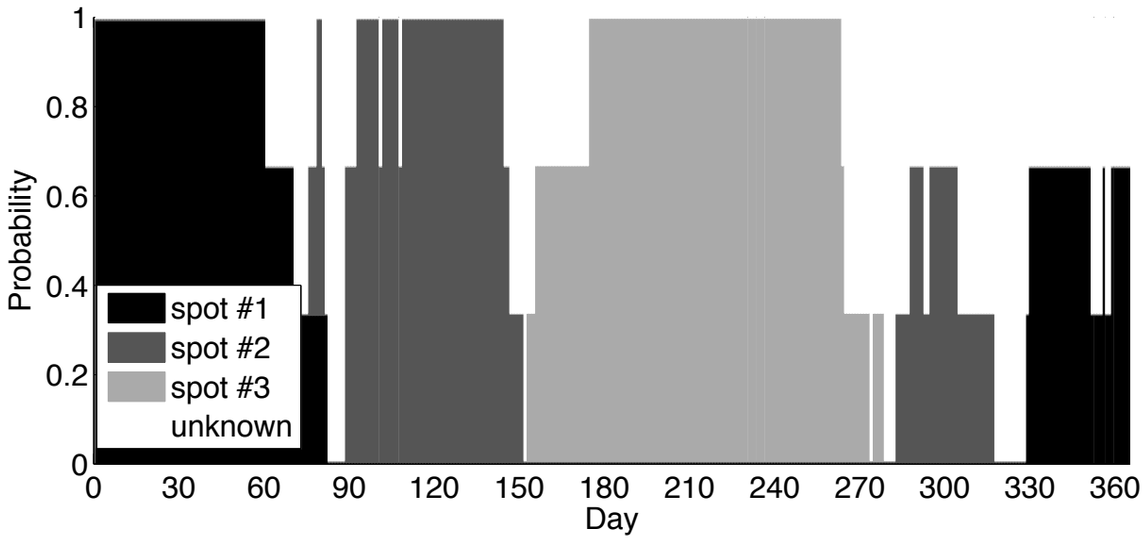
March May December

"This bald eagle stays in New York area (i.e., reference spot # 1) from December to March. In March, it flies to Great Lakes area (i.e., reference spot #2) and stays there until the end of May. It flies to Quebec area (i.e., reference spot #3) in the summer and stays there until late September. Then it flies back to Great Lake again staying there from mid October to mid November and goes back to New York in December."

A real bald eagle:
Step 2 Summarize behaviors

- Segment movement by 363 days
- "year"s are clustered into one cluster



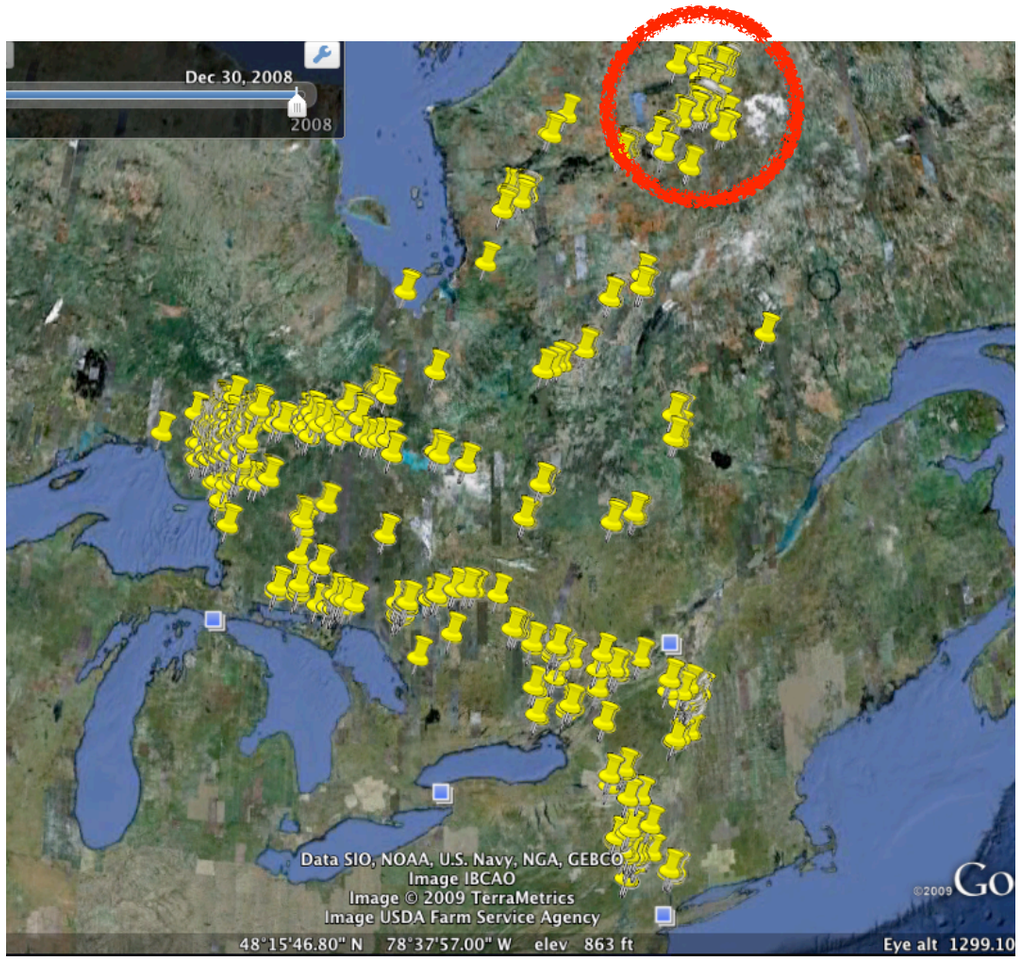


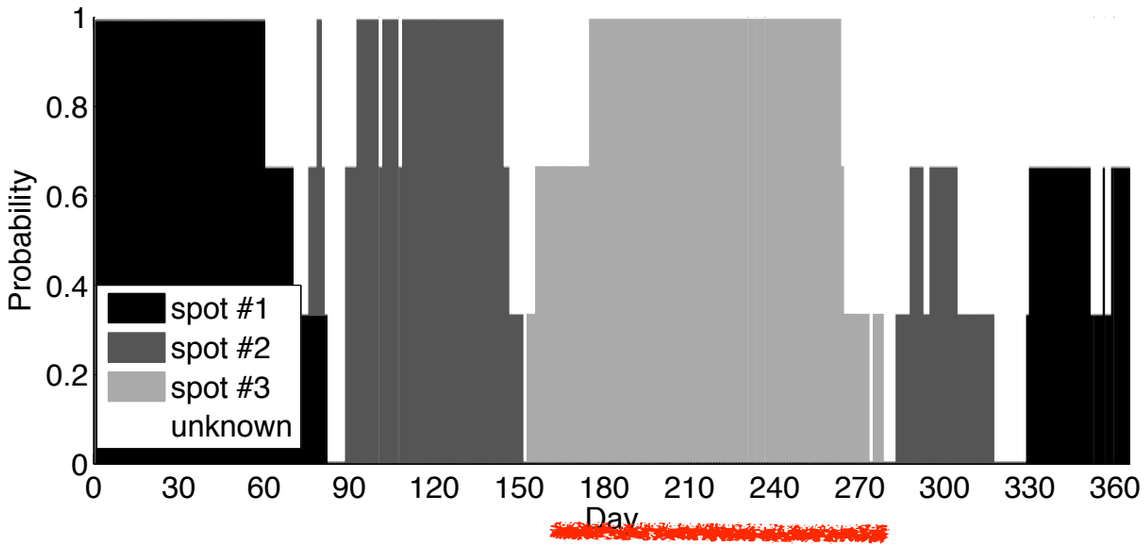
March May December

"This bald eagle stays in New York area (i.e., reference spot # 1) from December to March. In March, it flies to Great Lakes area (i.e., reference spot #2) and stays there until the end of May. It flies to Quebec area (i.e., reference spot #3) in the summer and stays there until late September. Then it flies back to Great Lake again staying there from mid October to mid November and goes back to New York in December."

A real bald eagle:
Step 2 Summarize behaviors

- Segment movement by 363 days
- "year"s are clustered into one cluster

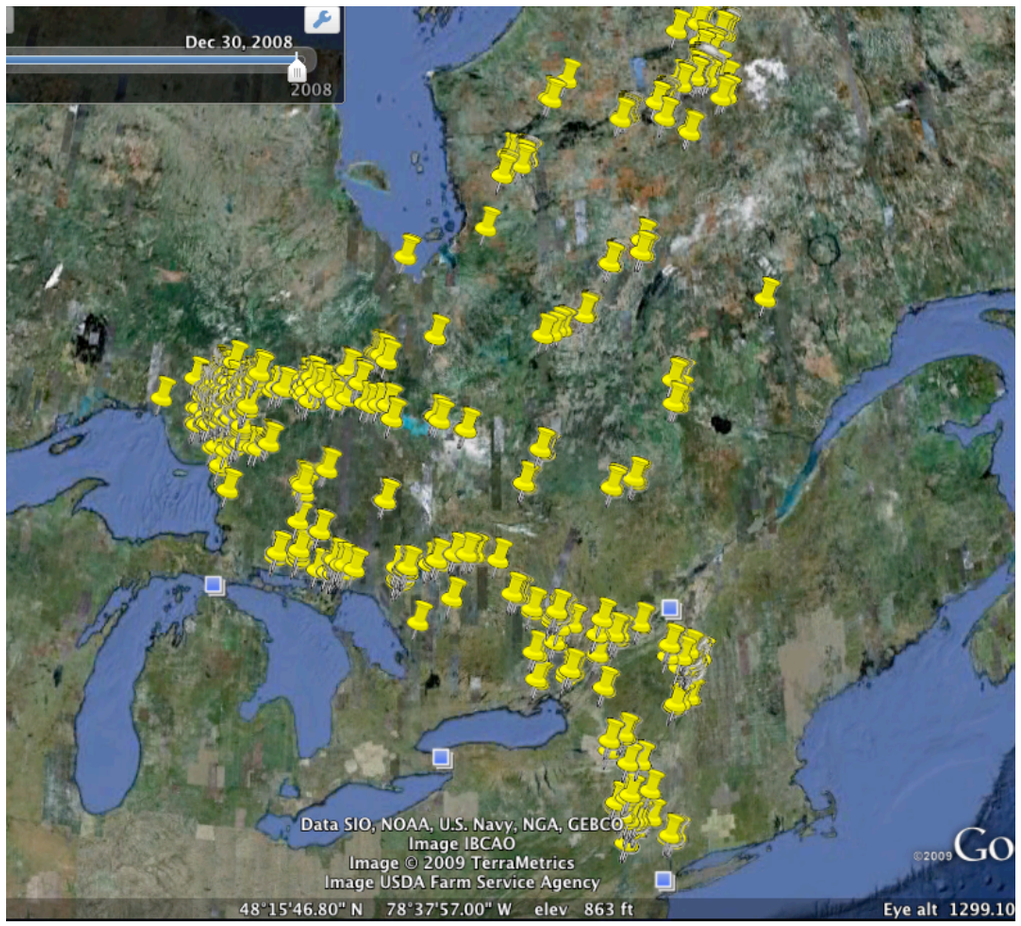


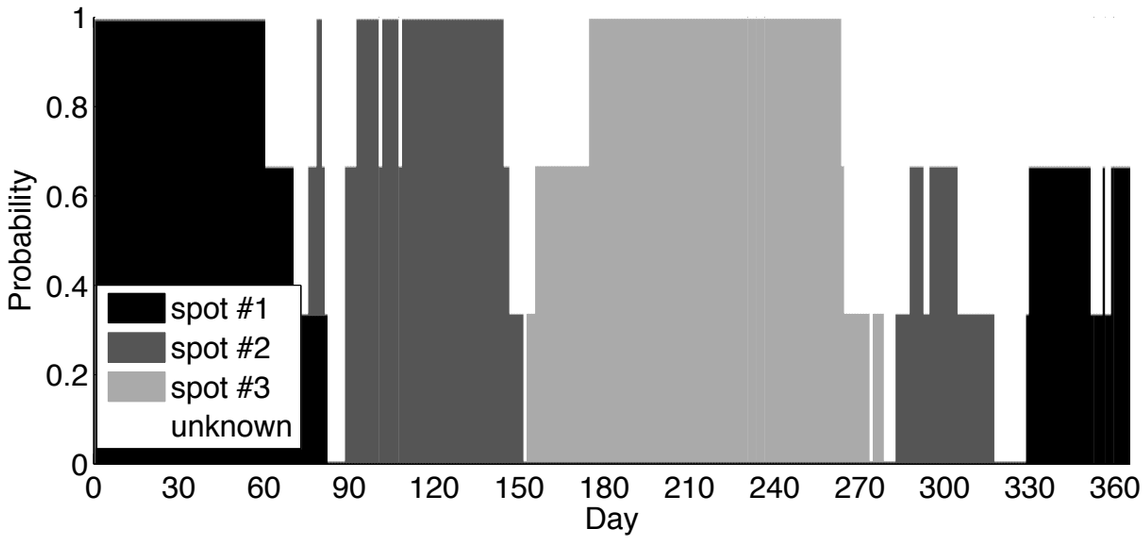


"This bald eagle stays in New York area (i.e., reference spot # 1) from December to March. In March, it flies to Great Lakes area (i.e., reference spot #2) and stays there until the end of May. It flies to Quebec area (i.e., reference spot #3) in the summer and stays there until late September. Then it flies back to Great Lake again staying there from mid October to mid November and goes back to New York in December."

A real bald eagle:
Step 2 Summarize behaviors

- Segment movement by 363 days
- "year"s are clustered into one cluster

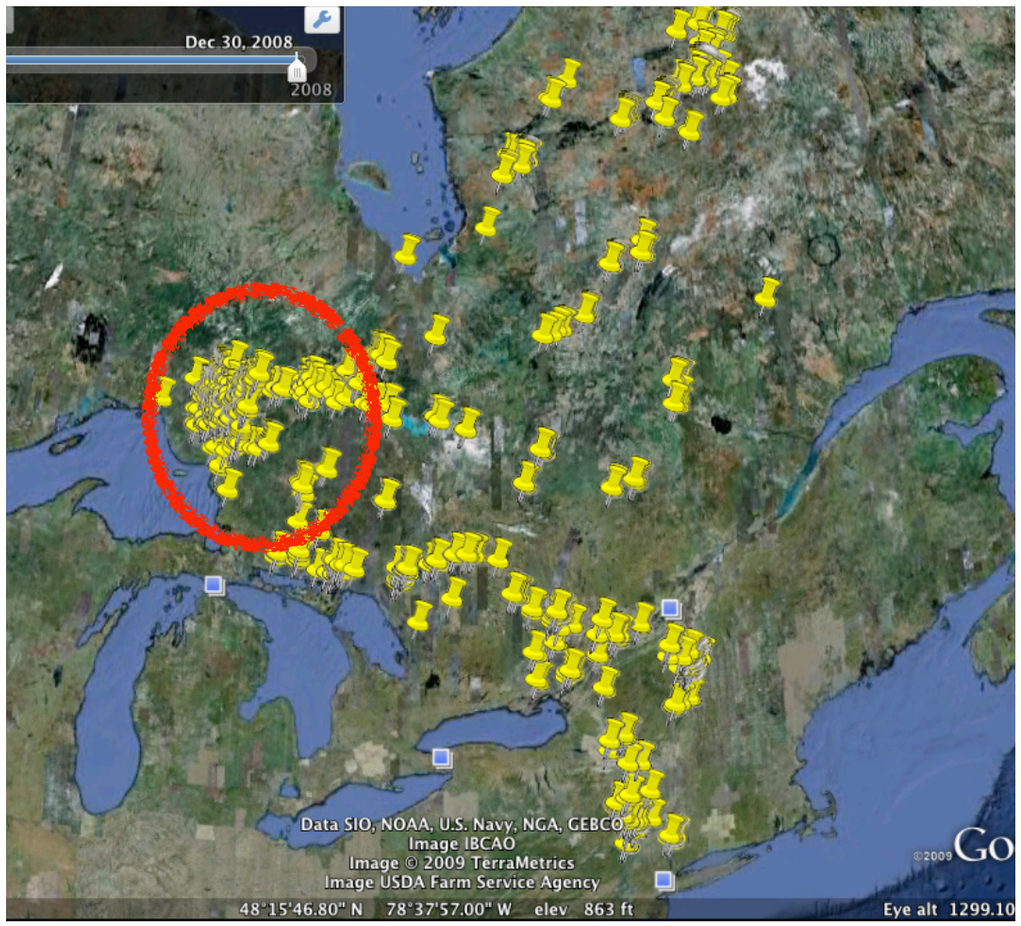


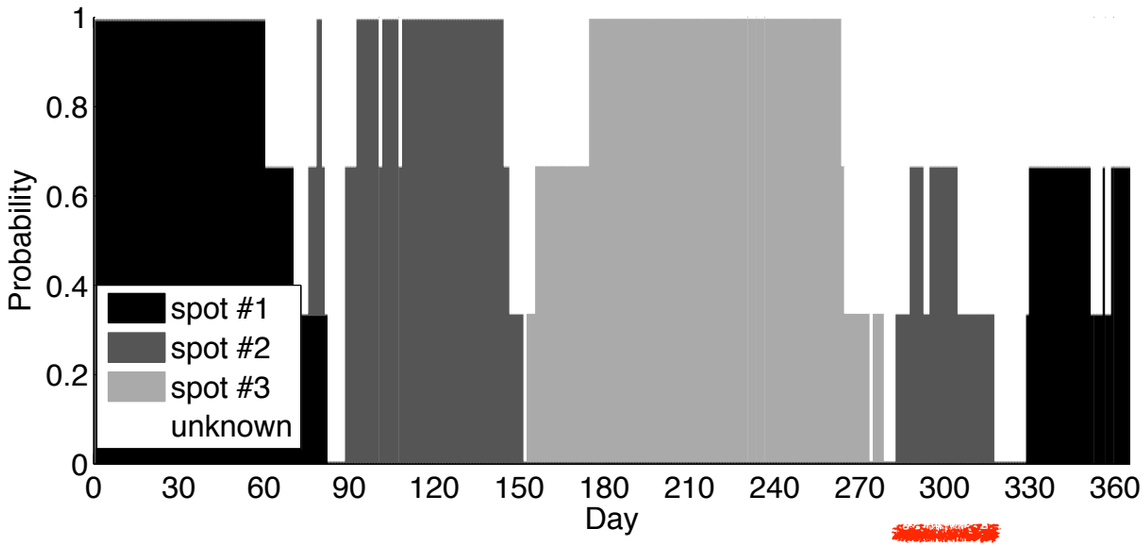


"This bald eagle stays in New York area (i.e., reference spot # 1) from December to March. In March, it flies to Great Lakes area (i.e., reference spot #2) and stays there until the end of May. It flies to Quebec area (i.e., reference spot #3) in the summer and stays there until late September. Then it flies back to Great Lake again staying there from mid October to mid November and goes back to New York in December."

A real bald eagle:
Step 2 Summarize behaviors

- Segment movement by 363 days
- "year"s are clustered into one cluster

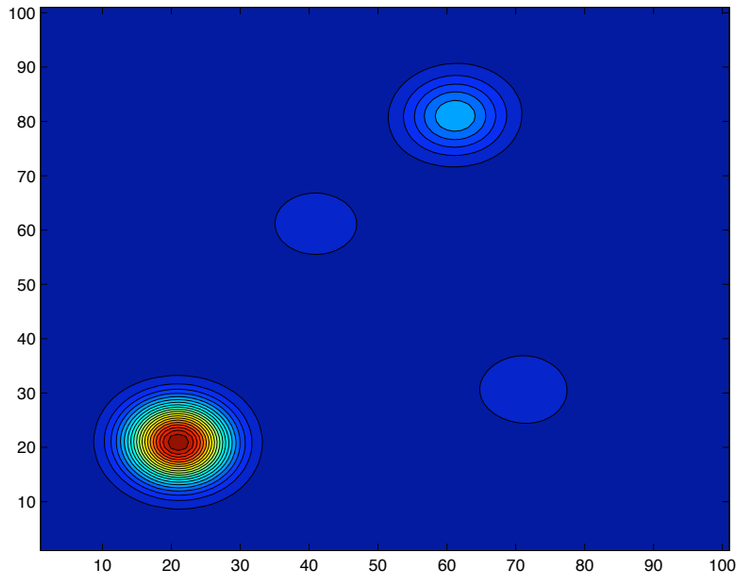




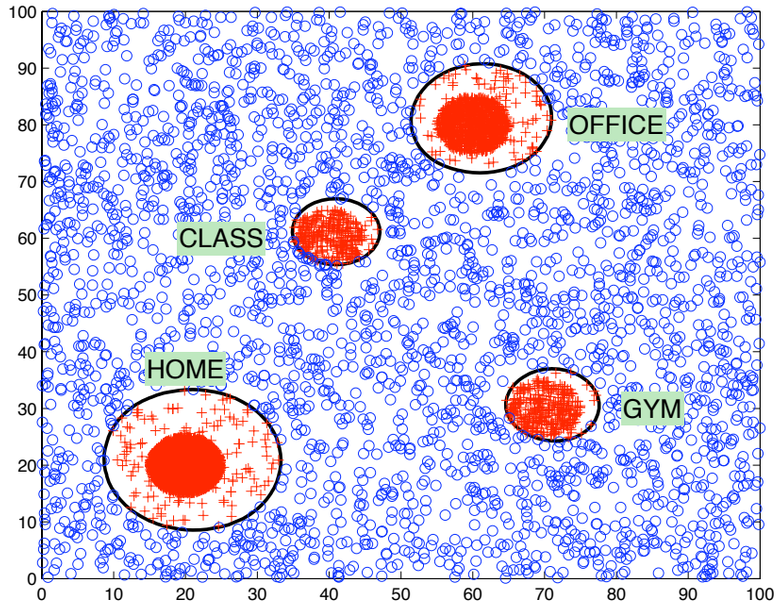
"This bald eagle stays in New York area (i.e., reference spot # 1) from December to March. In March, it flies to Great Lakes area (i.e., reference spot #2) and stays there until the end of May. It flies to Quebec area (i.e., reference spot #3) in the summer and stays there until late September. Then it flies back to Great Lake again staying there from mid October to mid November and goes back to New York in December."

Synthetic human movement:
Step 1 Detect periods

synthetic data:
20:00~8:00 at home
9:00~14:00 at office on weekdays
15:00~17:00 at gym on Tues. &Thurs.
15:00~17:00 at class on Mon., Wed., & Fri.



density map



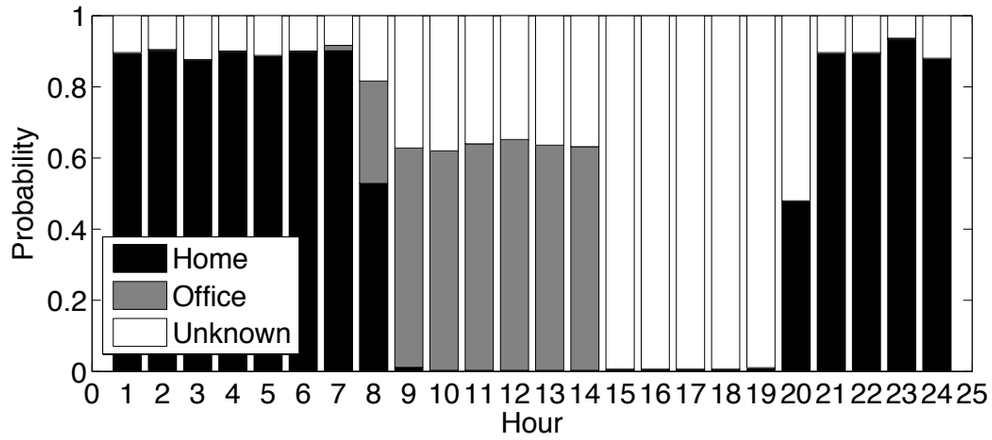
reference spots

Period detected using Fourier transform and autocorrelation:

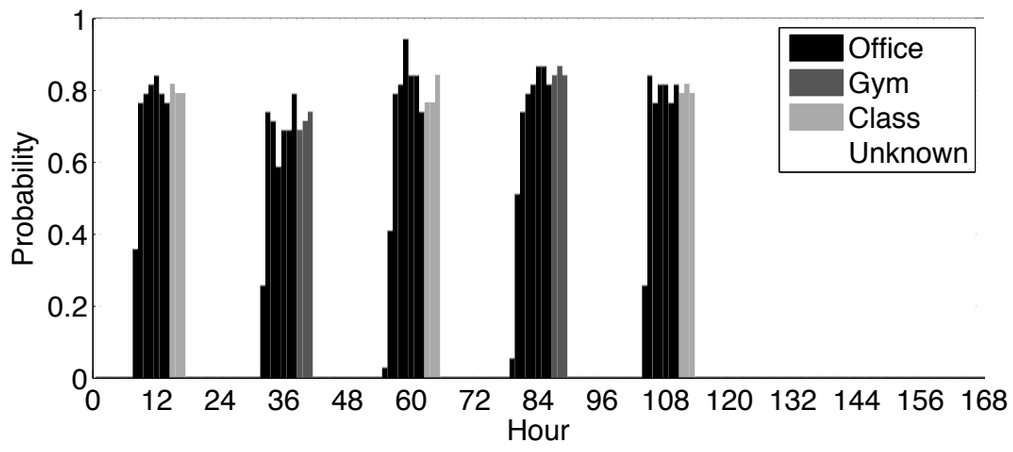
Obs. Spot	Home	Office	Gym	Class
Periods (hours)	24	24, 168	168	168

Synthetic human movement:
Step 2 Summarize behaviors

- Segment movement by **day** and **week** separately.
- Segments are clustered into one behavior.



daily periodic behavior



weekly periodic behavior

Outline

- Motivation
- Previous Method
- Periodica
- Experiment
- Summary

Summary

- We propose the idea of reference spots to detect periods.
  - Reference spots are detected using density-based method.
  - Periods are detect using FFT and auto-correlation.
- Periodic behaviors is modeled as a probabilistic matrix.
  - Behaviors are summarized via clustering.
  - Representation error is used to determine the number of behaviors.

Thanks!